A Modelling Attack Resistant Low Overhead Memristive Physical Unclonable Function

Xiaohan Yang*, Saurabh Khandelwal*, Aiqi Jiang+, and Abusaleh Jabir*

*School of ECM, Oxford Brookes University, UK. E-mail: {15058456, skhandelwal, ajabir}@brookes.ac.uk
+School of EECS, Queen Mary University of London, UK. E-mail: a.jiang@qmul.ac.uk

***Abstract*—Memristors are finding applications in memory, logic, neuromorphic systems, and data security. To this end, we leverage the non-linear behaviour of memristors to devise a low overhead physical unclonable function using a memristive chaos circuit in conjunction with a non-linear memristive encoder. We demonstrate the effectiveness of this architecture in Challenge-Response-Pair based authentication, and for its physical uncloneability. This architecture is highly versatile and can be implemented with a single encoder or a number of encoders running in parallel, each one with its own merit, for extending the sizes of CRPs. To demonstrate its effectiveness, we subject the architecture to machine learning based modelling attacks e.g. Logistic Regression, Support Vector Machines, Random Forest, as well as Artificial Neural Network classifiers. We found out that the proposed PUF architecture provides better resistance to such attacks, even for smaller bit sizes and at reduced overheads.**

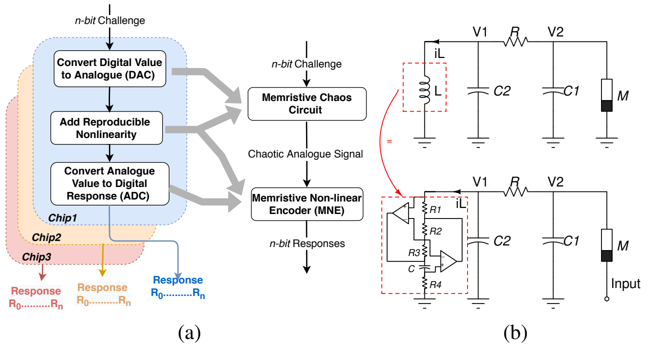

n-bit Challenge
Convert Digital Value to Analogue (DAC)
Add Reproducible Nonlinearity
Convert Analogue Value to Digital Response (ADC)
Chip1
Chip2
Chip3
Response R_0..........R_n
Response R_0..........R_n
Response R_0..........R_n
n-bit Challenge
Memristive Chaos Circuit
Chaotic Analogue Signal
Memristive Non-linear Encoder (MNE)
n-bit Responses
(a)
V1
R
V2
iL
L
C2
C1
M
=
V1
R
V2
iL
R1
R2
R3
C
R4
C2
C1
M
Input
(b)

Fig. 1: (a) Block diagram of the proposed architecture; (b) Memristive chaos circuits [9].

I. Introduction

A Physical Unclonable Function (PUF) is a promising low overhead security primitive useful for a number of authentication applications, e.g. in preventing integrated circuit piracy, cloning, and counterfeiting [1]. A PUF produces different unique outputs for the same input, when implemented on different devices. In Challenge-Response-Pair (CRP) based authentications, PUFs implemented on different chips produce unpredictable responses corresponding to the same challenges. Memristors, owing to their unique properties, are becoming popular in PUF designs [2]–[5]. A memristor stands apart from other fundamental circuit elements because of its non-volatility in the absence of a power supply. In this paper, we exploit this property for non-linearly encoding data and locking it to a particular device.

As size of the devices shrink, characteristics of memristors are affected by process variations. These characteristics, which are highly non-linear and pronounced, are fundamental behind its use as PUFs [2]–[5]. The main criticism and drawback of most of the existing PUFs, such as delay based arbiter PUFs, is that the mapping from challenges to responses shows a certain degree of linearity, which complex machine learning algorithms can figure out. For example, for 64-bit arbiter PUFs machine learning can achieve 99% prediction rate [6]. Additionally, for these architectures many CRPs lack uniqueness. To mitigate these usually a large number of challenge/response bits with many stages of iterative hardware, at the cost of significant overhead and power consumption, is required for a proper PUF [5], [7].

In contrast, we propose a novel architecture that leverages the non-linear behaviour of the memristors for realising high performance low overhead PUFs. It is based on a low overhead memristive chaos circuit in conjunction with a non-linear memristive encoder. The idea is to use the chaos circuits to non-linearly generate unique analogue values from different digital challenges and then use a non-linear memristive encoder to non-linearly convert these analogue values back to unique digital codes (responses), which requires much lower overhead and provides better attack resistance. Hence, the proposed architecture can be very useful in applications such as chip tagging/identification as well as for preventing unauthorised fabrications and authentication, e.g. via CRPs [5], [7], [8].

II. Proposed Architecture

Fig. 1a shows the block diagram of the proposed PUF architecture. The general idea is to convert the digital challenges to analogue signals, add non-linearity to it, and then convert the result back to digital responses. Here, the different chips are assumed to produce different responses for the same challenges owing to process variability. We achieve this with a low overhead memristive chaos circuit in conjunction with a Memristive Non-Linear Encoder (MNE). We show that, due to unpredictable non-linearity and process variability, the proposed architecture provides unpredictable CRPs as well as good PUF.

978-1-7281-9457-8/20/$31.00 ©2020 IEEE

A. Chaos Circuit

A reproducible memristive chaos circuit [9] is used for converting an n-bit digital challenge to a unique analogue value. The original chaos circuit constitutes a linear resistor, an inductor, two capacitors, and a nonlinear memristor. However, the inductor is large in size and difficult to scale and thus it is replaced by two op-amps [10] as shown by the equivalent circuit in Fig. 1b.

The idea of adding non-linear memristive chaotic property is to 'pre-scramble' the challenge, but in a reproducible way as long as the initial conditions are satisfied. Any change in the input alters the behaviour of the circuit completely and this change can be accumulated by applying the challenge bits in series. Hence, the chaotic analogue signal produced by the circuit has much less predictable relationship with the input challenges. This makes it much harder for a machine to learn the co-relation between the inputs and the analogue voltage and thus makes it difficult to mount a machine learning based attacks on the CRPs.

B. Memristive Non-Linear Encoder

Once the challenges are converted to analogue voltages, we use a low complexity MNE to encode the analogue signals to digital responses. This exploits the non-linear movement of a memristor's barrier, which make it even harder for CRP predictions. Blue line in Fig. 2 shows the basic architecture.

A memristor can be programmed (tuned) by applying a programming voltage V_{prog} and can be read by applying a hold voltage V_{hold}. The amplitude of the programming voltage V_{prog} and its pulse width T_{prog} determines the shifting of the barrier within the memristor [11]. Let f_{enc} be the encoding frequency of the clock (clk) and $T_{\text{enc}} = T_{\text{prog}} + T_{\text{hold}} = 1/f_{\text{enc}}$ be the clock cycle. Voltages V_{prog} and V_{hold} are alternatively applied during T_{prog} and T_{hold} respectively. V_{prog} is adjusted to be sufficiently high so that V_{W} appears across the memristor even with the load R_{L}. The memristor is initialised to R_{off} and by repeatedly applying V_{prog}, voltage V_b increases towards V_a. During each T_{prog}, V_{prog} shifts the barrier of the memristor from the R_{off} region towards the R_{on} region by a small amount, but non-linearly, and during the following T_{hold}, V_b and V_a are compared. The barrier of the memristor can also be shifted from R_{on} to R_{off} except the fact that the polarity of the memristor needs to be switched. Meanwhile, an m-bit counter counts the number of programming pulses. The programming pulses represent the encoded value, which is non linear, corresponding to V_a. The control circuit disables the programming pulses when V_b exceeds V_a.

The counter, and hence the encoded value, can be of any number of bits. For example, a 1-bit counter flips between 0 and 1 during the encoding process and provides a 1-bit encoding of V_a. This outcome depends on the non-linear shifting of the barrier within the device. This flexibility allows us to segment an n-bit response into k smaller bit responses and then combine the results.

C. Proposed PUF Architecture

Inline with Fig. 1a, Fig. 2 shows the proposed architecture for CRP based authentication. The architecture generates unpredictably non-linear CRPs owing to the combined effects of the chaos circuit and the MNE. It also provides physical unclonability by virtue of its sensitivity to process and parametric variations. As revealed by our experimental results, the CRPs depend heavily on the physical parameters of a memristor such as its length, D. Any small variations in these are amplified by the MNE mechanism and results in wide variations in responses for the same challenges (Fig. 3a).

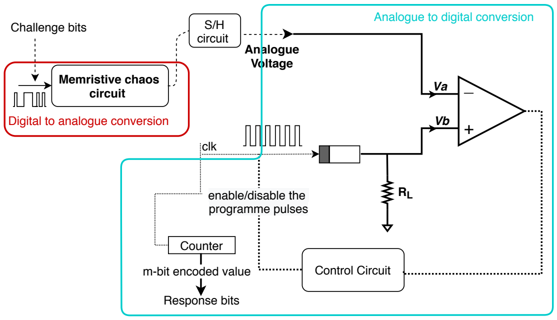


Fig. 2: Proposed architecture for CRP based authentication.

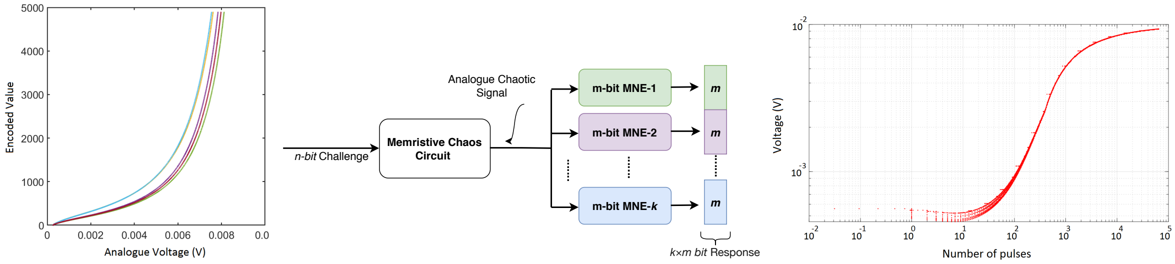
Fig. 3b shows the proposed PUF architecture with multiple MNEs placed in parallel. These paralleled MNEs produce different m-bit responses for the same voltage (challenge) by virtue of process variability. Each m-bit response is concatenated to generate a new unique $k \times m$ bit response. Dividing a single m-bit response into k smaller instances can be useful, e.g. when m is very large.

The proposed architecture also reduces the effects of quantisation error and improves uniqueness. As shown in Fig. 3c, due to quantisation error some analogue values (challenges) produced by the chaos circuit are mapped to the same encoded value (response) in a single chip. Concatenating responses from different MNEs lowers this effect.

III. Results and Discussions

For the experimental results, the memristors were coded in Verilog-A based on the model and the parameters given in [11]. The systems were designed and simulated in Cadence Virtuoso. We assumed V_{prog}= 250mV, V_{hold}= 50mV, T_{prog}= 2.5ns, R_{L} = 1KΩ, R_{on}=500Ω, R_{off}=200KΩ, D=3nm, V_{on}=−0.2V and V_{off}=0.02V.

a) Non-linear Memristive Encoder: The MNE architecture presented by blue line in Fig. 2 inherently provides non-linear encoding as shown in Fig. 3a. This figure also shows that a small variation in the physical parameters of memristor results in different analogue-to-digital transfer characteristics. As we see the behaviour of the MNE is non-linear throughout, i.e. it is non-linear for specific V_{prog} or T_{prog} and also for their differences. The MNEs also showed wide variations in responses for the same challenges under process variations when used as a CRP generator for authentication.



(a) (b) (c)

Fig. 3: (a) Effects of varying process parameter D by 5%, while all other parameters are fixed; (b) Proposed PUF architecture with multiple MNEs; (c) Effects of the quantisation error.

TABLE I: Hardware overhead comparison for memristor-based PUFs.

PUF Architecture	Ref. [2]	Ref. [3]	Ref. [4]	Ref. [12]	Proposed Architecture
Challenge size (bits)	n	n	n	n	n
Response size (bits)	r	r	r	r	$r = k \times m$
Memristors	$n \times r$	$n \times r$	$2n$	$4n \times r$	$k+1$
MOSFET switches	–	–	–	$4n \times r$	–
MUXes(2:1)	$3n+r$	–	$3n$	r	–
Resistors	–	–	$2n$	–	$k+5$
Amplifiers	r	r	–	–	2
Inverters	–	–	$2n$	$2n \times r$	–
n-bit decoder	–	1	–	–	–
Flip-flops	–	–	–	$2 \times r$	r
Capacitors	–	–	–	–	3
Comparator	–	–	–	–	k

b) Hardware Overhead: Table I shows the hardware overhead comparison for proposed PUF architecture and existing memristor-based PUFs [2]–[4], [12]. Table I is originally extracted from [12], which exclude the timing and control circuit for all of the techniques. From the results, the hardware requirement in [2]–[4], [12] are much higher compared to the proposed PUF design. The proposed architecture is also flexible in terms of hardware overhead as the value of k and m can be set as per the requirement of the specific design.

c) Performance: Various metrics to determine the quality of PUF have been proposed in [13]. Table II shows the performance of the proposed architecture compared to existing architectures [5], [7], [8].

Uniqueness: Uniqueness determines the ability of different PUFs to produce unique responses to the same challenge. It is defined as $\frac{2}{k(k-1)}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{\mathrm{HD}(R_i,R_j)}{n} \times 100\%$, where k is the total number of PUF instances and n represents the total number of bits in the response of a PUF instance; $\mathrm{HD}(R_i,R_j)$ is the hamming distance between two bit-strings R_i and R_j. Ideally, the uniqueness should be 50%. For the proposed PUF architecture, when tested with $1,000$ different PUF instances, it is 49.92%.

Uniformity: It is a measure of the distributions of 1s and 0s in an instance of a PUF. Even distribution ensures a strong security key. Uniformity for Chip-i is calculated as $\frac{1}{n}\sum_{l=1}^{n} r_{i,l} \times 100\%$, where $r_{i,l}$ is the lth binary bit of an n-bit response [13]. The ideal uniformity is 50%. The uniformity of the proposed architecture, when tested with 64-bit 60,000 different CRPs, is 56.33%.

Bit Aliasing: It is a measure for the likelihood of different chips producing similar response bits. The ideal value is 50%. For Chip-i, it is calculated as $\frac{1}{k}\sum_{i=1}^{k} r_{i,l} \times 100\%$, where $r_{i,l}$ is the lth bit of an n-bit response and k is the total number of PUF chips [13]. The Bit Aliasing for the proposed architecture, when tested with 1,000 different PUFs, is 49%.

TABLE II: Performance comparison.

	Uniqueness	Uniformity	Aliasing
Ref. [5]	50.3%	53.8%	49.2%
Ref. [7]	49.8	50.1%	–
Ref. [8]	50%	50.69%	50%
Proposed PUF	49.92%	56.33%	49%

d) Modelling Attack Resistance: Most existing PUFs are susceptible to machine learning based modelling attacks [6]. Especially for delay based PUFs (e.g. Arbiter PUFs), modelling attack is extremely successful. The CRPs in the existing PUFs are either linearly separable or mathematically differentiable. In contrast, we show that our PUF architecture is capable of offering higher resistance to such attacks because of the non-linear stochastic properties provided by chaos and MNE circuits.

The modelling attacks we applied in this paper are under the environment of Python 3.6 with the packages Scikit-Learn 0.19.0 and Keras 2.2.4. Logistic Regression (LR), and Support Vector Machine (SVM) are the most widely used techniques for the modelling attacks on PUFs. In our case, to demonstrate the general machine learning attack resistance of the proposed design, we not only use SVM and LR, but also apply other techniques such as Decision Tree (DT) and Random Forest (RT) classifiers. In SVM attacks, we specifically chose Radial Basis Function (RBF) kernel to fit our nonlinear property [7]. We also tested attack resistance to Artificial Neural Network (ANN) based classifiers. To solve the non-linear problem, we use Multi-layer Perceptron (MLP) feed-forward network structure for classification. ANN based classifiers have been claimed to outperform traditional machine learning algorithms [8].

To improve the machine learning performance, in addition to the CRPs we also considered the voltage V_a (Fig. 2) as an additional feature. However, in reality, unlike existing

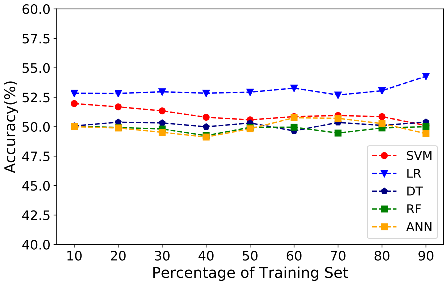


Fig. 4: Modelling attack results for different size of the training set.

TABLE III: Modelling attack resistance comparisons.

Class. Learner	Ref. [5] %Accuracy	Ref. [7] %Accuracy	Ref. [8] %Accuracy	Proposed PUF %Accuracy
LR	50	–	–	53.06
SVM	65.67	79	–	49.81
DT	–	–	–	50.65
RF	–	–	–	49.43
ANN	–	–	≈50	53.4

TABLE IV: Performance with and without MNE.

Classification Learner	64-bit Challenge with different bit size response (%Accuracy)					
	1-bit		2-bits		4-bits	
	LE	MNE	LE	MNE	LE	MNE
SVM	94.3	49.81	62.6	26.26	23.1	10.02
LR	93.7	53.06	64.5	30.12	23.1	13.86
DT	86.1	50.65	52.8	25.30	14.2	7.37
RF	93.6	49.43	61.1	24.5	18	6.64
ANN	93.3	53.4	63.2	24.1	23.11	4.6

iterative network based PUFs [5], [7], these types of features are much harder to obtain because of the sequential nature of the hardware and accessibility. Here, V_a is dependent on the previous voltage levels, which is much harder to track. However, without this feature the accuracy of all the classifiers drops below 50%.

All classifiers were trained with 60,000 CRPs. Fig. 4 presents the trends of accuracy, which randomly fluctuate around 50% rather than generally increase as the training set is extended. This indicates that the proposed PUF is able to effectively resist modelling attacks. We also compare the attack resistance of the proposed design with the existing PUFs which are subjected to same machine learning attacks. The results are summarised in Table III. In this table, the training sets considered by [5], [7], and [8] are 5,000, 50,000, and 38,000 respectively. The proposed design provides better resistance to the attacks compared to [5], [7]. It is worth mentioning that for the traditional arbiter PUFs, these attacks can reach almost 99% accuracy. For the ANN classifier, the accuracy of our technique is 53.4%. If we consider only the CRPs, without additional features, the accuracy drops below 50%. This is an improvement over [8] also, which did not seem to specify any additional feature for training.

To evaluate the effects of the MNE, we tested attack resistance without it, i.e. with a Linear Encoder (LE). The results appear in Table IV. Clearly, the accuracy are around 90% without the MNE, and drops to around 50% with the MNE. We also show here the effects of using higher bit counters and training with multi way classifiers. Clearly, the accuracy is drastically reducing for both, but with the MNE it is dropping further.

IV. Conclusions

In this paper we presented a novel concept on low overhead and high performance PUF realisation based on a non-linear decoding and encoding scheme. The scheme was tested with a low overhead memristive chaos block as the decoder in conjunction with versatile memristive non-linear encoders. The proposed architecture is sequential in nature and offers much lower overhead compared to existing techniques and higher flexibility in terms of hardware requirements. The combined non-linearity of the decoder and encoder offers excellent resistance to modelling attacks. Our results showed that the proposed architecture can outperform existing ones. Owing to its versatility, non-linearity, physical unclonability and lower overhead, we envisage that the proposed architecture will be attractive for diverse security, authentication, and trust applications.

References

[1] G. E. Suh and S. Devadas, "Physical unclonable functions for device authentication and secret key generation," in *Proceedings of the 44th Annual Design Automation Conference*, 2007, p. 9–14.

[2] J. Rajendran, G. S. Rose, R. Karri, and M. Potkonjak, "Nano-ppuf: A memristor-based security primitive," in *2012 IEEE Computer Society Annual Symposium on VLSI*, 2012, pp. 84–87.

[3] P. Koeberl, U. Kocabaş, and A. Sadeghi, "Memristor pufs: A new generation of memory-based physically unclonable functions," in *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*, 2013, pp. 428–431.

[4] G. S. Rose, N. McDonald, L.-K. Yan, B. Wysocki, and K. Xu, "Foundations of memristor based puf architectures," in *2013 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*. IEEE, 2013, pp. 52–57.

[5] J. Mathew, R. S. Chakraborty, D. P. Sahoo, Y. Yang, and D. K. Pradhan, "A novel memristor-based hardware security primitive," *ACM Trans. Embed. Comput. Syst.*, vol. 14, no. 3, Apr. 2015.

[6] L. Chen, "A Framework to Enhance Security of Physically Unclonable Functions Using Chaotic Circuits," *Phys. Lett. A*, vol. 382, no. 18, pp. 1195–1201, 2018.

[7] A. Vijayakumar and S. Kundu, "A Novel Modeling Attack Resistant PUF Design Based on Non-Linear Voltage Transfer Characteristics," in *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, 2015, pp. 653–658.

[8] E. I. Vatajelu, G. Di Natale, M. S. Mispan, and B. Halak, "On the encryption of the challenge in physically unclonable functions," in *2019 IEEE 25th International Symposium on On-Line Testing and Robust System Design (IOLTS)*. IEEE, 2019, pp. 115–120.

[9] L. O. Chua, "The Genesis of Chua's Circuit," *AEU-Int J Electron C.*, vol. 46, no. 4, pp. 251–257, 1992.

[10] V. Siderskiy, "Chua's Circuits," http://www.chuacircuits.com.

[11] S. Kvatinsky, M. Ramadan, E. G. Friedman, and A. Kolodny, "Vteam: A General Model for Voltage-Controlled Memristors," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 62, no. 8, pp. 786–790, Aug 2015.

[12] J. Mathew, R. S. Chakraborty, D. P. Sahoo, Y. Yang, and D. K. Pradhan, "A novel memristor based physically unclonable function," *Integration, the VLSI Journal*, vol. 51, pp. 37 – 45, 2015.

[13] A. Maiti, V. Gunreddy, and P. Schaumont, "A systematic method to evaluate and compare the performance of physical unclonable functions," in *Embedded systems design with FPGAs*. Springer, 2013, pp. 245–267.